header for SPIE use

Apr 98

An Evaluation of SAR ATR Algorithm Performance Sensitivity to MSTAR Extended Operating Conditions

John C. Mossing[a], Timothy D. Ross[b], and Jeff Bradley[a]

[a]Sverdrup Technology, Inc. Advanced Systems Group 4200 Col. Glenn Highway, Suite 500, Beavercreek, OH 45431
[b]Air Force Research Laboratory, AFRL/SNA, Wright-Patterson AFB, OH 45433

ABSTRACT

Testing a SAR Automatic Target Recognition (ATR) algorithm at or very near its training conditions often yields near perfect results as we commonly see in the literature. This paper describes a series of experiments near and not so near to ATR algorithm training conditions. Experiments are setup to isolate individual Extended Operating Conditions (EOCs) and performance is reported at these points. Additional experiments are setup to isolate specific combinations of EOCs and the SAR ATR algorithm's performance is measured here also. The experiments presented here are a by-product of a DARPA/AFRL Moving and Stationary Target Acquisition and Recognition (MSTAR) program evaluation conducted in November of 1997. Although the tests conducted here are in the domain of EOCs, these tests do not encompass the "real world" (i.e., what you might see on the battlefield) problem. In addition to performance results this paper describes an evaluation methodology including the Extended Operating Condition concept, as well as, data; algorithm; and figures of merit. In summary, this paper highlights the sensitivity that a baseline Mean Squared Error (MSE) ATR algorithm has to various operating conditions both near and varying degrees away from the training conditions.

Keywords: Performance Sensitivity, Automatic Target Recognition, Synthetic Aperture Radar, Extended Operating Conditions, Algorithm Evaluation, Mean Square Error Classifier, MSTAR

1. INTRODUCTION

Synthetic Aperture Radar (SAR) air-to-ground images are collected by various platforms (e.g., the U2, Global Hawk, or F-15E) for various purposes (e.g., reconnaissance or targeting). The collection capacity for such imagery is growing rapidly, and along with that growth is the expanding need for computer-aided or automated exploitation of SAR images. One aspect of the aided/automated exploitation is automatic target recognition (ATR). An ATR algorithm finds target like regions (regions-of-interest (ROIs)) within a SAR image and computes a class (e.g., T72, BTR70, ...) for each ROI. ATR algorithm development typically involves some "training" with example images of known objects.

ATR evaluations are an important basis for technical and programmatic decision making. Technical decisions might include the choice between two alternative approaches to some sub-system or the identification of weak-links within a system. Programmatic decisions might involve competitive down-selects or technology transitions to users. The evaluation of any "trained" system has its challenges, but the evaluation of ATRs seems to be especially problematic because of the constraints on the available data compared to the complexity of the problem.

This paper makes two main contributions. An evaluation methodology is introduced and demonstrated that we believe will lead to results that more reliably support the decision making process. In demonstrating that methodology, the sensitivity of SAR ATRs' to specific (although certainly not all) factors that make the problem complex is reported. The result, we hope, is a better understanding of the SAR ATR problem.

INTERNET DOCUMENT INFORMATION FORM

A . Report Title: An Evaluation of SAR ATR Algorithm Performance Sensitivity to MSTAR Extended Operating Conditions

B. DATE Report Downloaded From the Internet 11/20/98

C. Report's Point of Contact: (Name, Organization, Address, Office Symbol, & Ph #): Air Force Research Laboratory
AFRL/SNA
ATTN: Timothy D. Ross
Wright-Patterson AFB, OH 45433

D. Currently Applicable Classification Level: Unclassified

E. Distribution Statement A: Approved for Public Release

F. The foregoing information was compiled and provided by:
DTIC-OCA, Initials: VM_ **Preparation Date:** 11/20/98__

The foregoing information should exactly correspond to the Title, Report Number, and the Date on the accompanying report document. If there are mismatches, or other questions, contact the above OCA Representative for resolution.

2. BACKGROUND

2.1 MSTAR Program Overview

The MSTAR Program Area is focused on the development, integration and evaluation of advanced automatic target recognition (ATR) systems. These systems are capable of high-performance identification of tactical and strategic targets in synthetic aperture radar (SAR) imagery. Such technology is absolutely necessary to achieve dominant battlefield awareness in the face of huge volumes of surveillance imagery. The MSTAR program will provide these automated algorithms and tools for insertion into future advanced technology demonstrations.

The primary goals of the MSTAR Program are to: 1) Develop, integrate and evaluate in the lab a robust and reliable system for the recognition of a twenty target set containing high value tactical and strategic targets: 2) Develop integrated approaches for recognizing targets under variable sensor and deployment conditions: sensor squint, depression and aspect angles; target articulation, configuration, shadow obscuration, terrain layover, and camouflage: 3) Develop and use a modular model-driven ATR architecture to facilitate component evaluation, subsystem re-use and system adaptation to new targets and mission conditions.

The MSTAR program is currently in development. In November of 1997 MSTAR completed the second year of a 3 year base program. The experiments presented in the following sections are a by-product of a November '97 evaluation. An associated team of module developers and a system integrator is developing the MSTAR system under a distributed, collaborative development strategy.[1]

2.2 MSTAR Evaluation

Although this paper focuses on a "baseline" Automatic Target Recognition (ATR) algorithm's sensitivity to extended operating conditions, it is the by-product of an independent evaluation of a more complex ATR system. The baseline algorithm is presented in section 3.2.2. The MSTAR program evaluations are conducted by the Air Force Research Lab (AFRL)/Sverdrup Technology (SvT) Performance Evaluation (PE) team. The MSTAR evaluation is an "in the loop" assessment and feedback process to the Program Office and algorithm developers. This process which is depicted in Figure 1, contributes to the system design, program continuance decision, and technology investment/transition directions. Evaluation activities include "test" planning, data sequestration, evaluation infrastructure/tool development, installing algorithms, "test" execution, analysis, and reporting of results. "Tests" fall into two categories "System Analysis Experiments" and "Evaluations". System Analysis Experiments are conducted with Unsequestered data that the algorithm developers have access to. Evaluations are conducted on Sequestered data that the algorithm developers do not have access to.

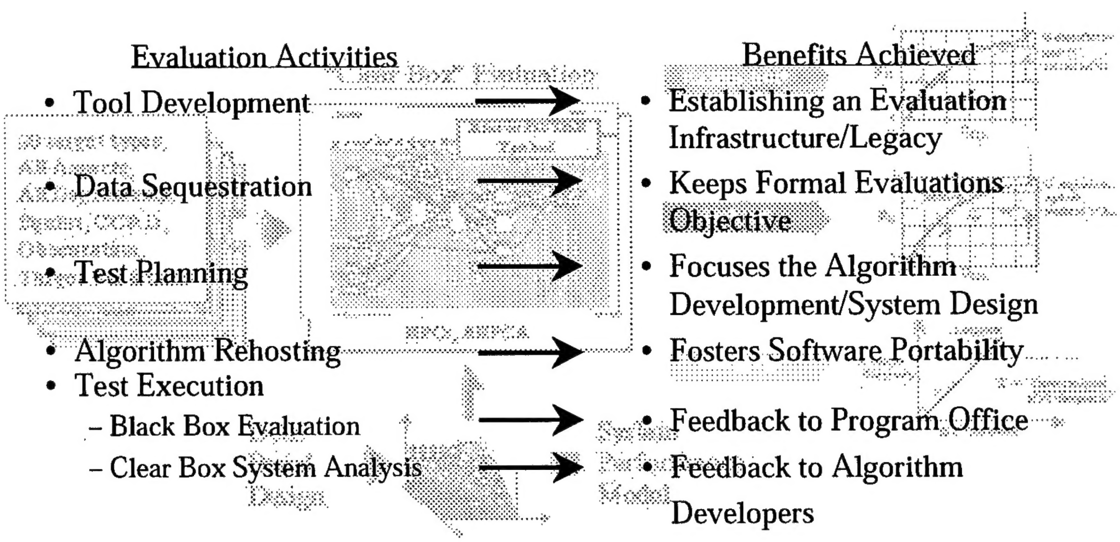


Figure 1. MSTAR Evaluation Activities and Benefits.

The Performance Evaluation team reports to the DARPA/AFRL Program Office, The effort is conducted on site at AFRL, WPAFB. Evaluations are conducted periodically throughout the development cycle of the program. These evaluations are goal oriented which focus development efforts.

3. EVALUATION CONCEPTS AND COMPONENTS

3.1 Concept: Extended Operating Conditions

In general, for an ATR algorithm to be successful it must be able to maintain high probability of detection (Pd) and probability of identification (Pid) rates while maintaining a low false alarm rate (FAR) over a variety of *Extending Operating Conditions*. This section attempts to define what the authors mean by Extended Operating Conditions. We start with a scenario of how many ATR algorithms are developed and tested. Collected measured data is commonly needed to develop an ATR. Often a specific data collection is designed to support the development of that particular ATR or data is leveraged off a previous collection with similar goals. Once a data set is available for ATR development, typically it is divided into "train" and "test" sets. Many methods are used to decide which data are used for training and which data are used for testing, some are as simple as placing every other collected sample into the train set with the remaining samples are placed in the test set. In this type of odd/even train/test data partitioning scenario, the ATR algorithm is tested with the test data, which has properties nearly identical to the training conditions. This type of testing is highly discouraged since the results obtained from it would be overly optimistic. The above is an example of what Extended Operating Condition testing is *NOT*. In fact, in the above scenario the testing OCs are as close as one could get to the train conditions without utilizing training data in the testing process. An Extended Operating Condition (EOC) is defined to be at an Operating Condition (OC) "away" from the trained condition. The MSTAR EOCs are presented in section 4.[2,3]. EOC testing is crucial for determining if an ATR algorithm is ready to be fielded.

3.2 Components

Three important ingredients needed to conduct an ATR evaluation are the data, the algorithms, and the performance measures. To conduct an efficient evaluation, obtaining the benefits depicted in Figure 1, it is extremely important that each of these ingredients (data, algorithms, and performance measures) are well understood prior to conducting the evaluation. These components are shown in Figure 2. Sections 3.1 through 3.3 describe them in more detail.

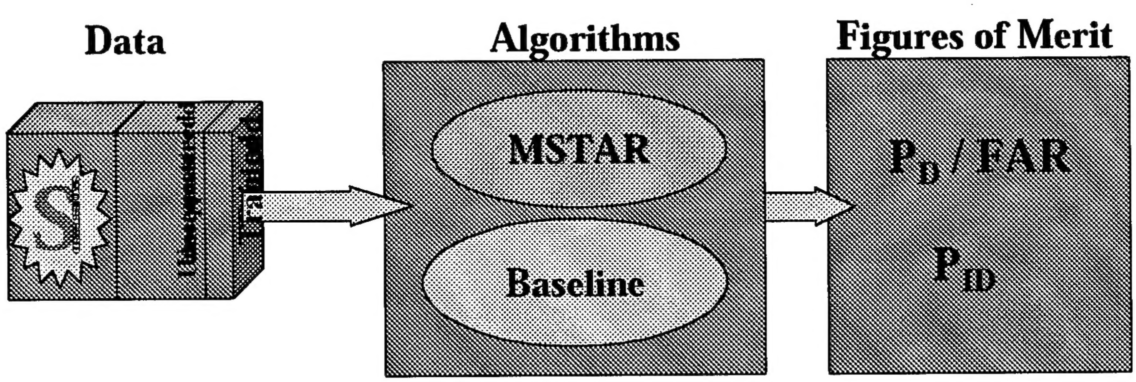


Figure 2. SAR ATR Evaluation Components data, algorithms and performance measures.

3.2.1 The MSTAR Data

To date, the MSTAR program has sponsored three data collections. These collections have resulted in hundreds of thousands of target samples and approximately one hundred square kilometers of clutter data. All of the SAR data was collected at X band, HH polarization, 1'x1' resolution. Figure 3, depicts the "states" of the MSTAR clutter and target scenes. Clutter data was collected at three depressions (15°, 30° and 45°). It was then categorized into three classes, Rural, Sparsely Built-Up, and Built-Up. The target scenes collected were well truthed, contained approximately 25 target types and controlled variations of: target component articulations (guns, turrets, hatches, etc.), revetments, squint, depression, 0-360° target aspects, multiple

backgrounds and version variants. The AFRL/SvT Evaluation team has written an "MSTAR Data Handbook for Experiment Planning". The handbook is over 100 pages and is the result of a rather extensive study of all of the MSTAR data. This handbook can be made available by contacting the authors. Once the data was well sorted, sequestration plans were developed, allowing the algorithm developers access to only portions of the data. For more information on the MSTAR data visit the AFRL Data Teams' web site[4].

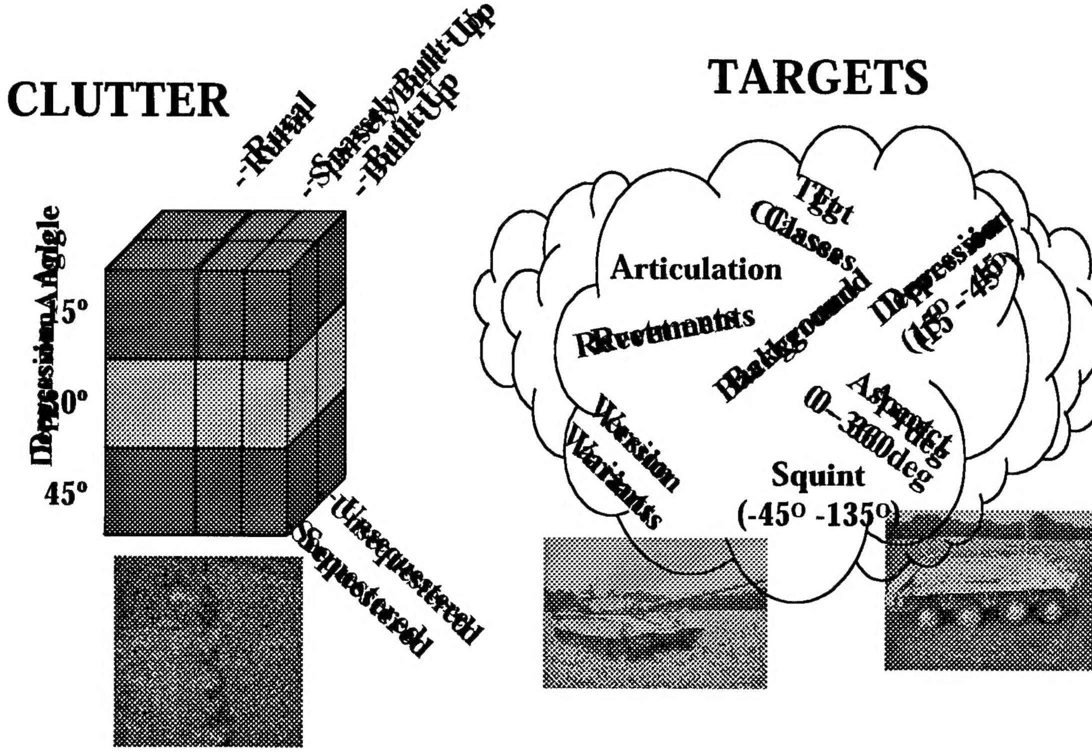


Figure 3. MSTAR data used for testing. All testing discussed in this paper utilized measured MSTAR data. Clutter data was collected at 3 depressions and categorized into three classes. Target scenes contained many EOC states.

3.2.2 Baseline SAR Algorithm Attributes

The baseline SAR algorithm used in this study consisted of three parts: CFAR filter, a target-like/nontarget-like Discriminator, and a Mean Squared Error Classifier. These components and their functions are depicted in Figure 4. The CFAR filter takes as input a full-scene SAR image. The CFAR's job is to locate the Regions of Interests (ROIs) which contain areas of "potential targets". The criteria for something to be a potential target are: a) it be an area that contains some bright pixels and b) and that the cluster of these bright pixels be within two predetermined sized bounding boxes which determine the minimum and maximum size of the targets of interest. The output of the CFAR filter is a row, column location in the full-scene image of the center of the ROI and a CFAR test statistic. The second of the three stages of this ATR Algorithm is the Discriminator. The Discriminator inputs are ROIs as determined by the CFAR. The Discriminator's job is to reject ROIs which aren't "target-like". It does this by training a series of texture, shape, and contrast features. The last of the three stages is the Mean Squared Error template matcher. The MSE's job is to produce a target identification declaration for target ROIs and to reject non-target (clutter) ROIs as non-targets. This stage of the algorithm is also trained. For tests presented in this paper the algorithm was trained across 10 (and 15) target classes. MSE Training for each target class consisted of approximately 225 test chips covering 0°-360° in target aspect at a depression angle of approximately 30°. Training was limited to a single Serial Number (SN) for each target type. A target set was defined which contained many Extended Operating Condition ROIs. These ROIs were sent through the algorithm and an operating point was set such that 90% of the ROIs were detected by the system. Once this operating point was set it was not changed. The algorithm was then tested across many individual and combination of EOC target sets.

Although more complex than the above-mentioned algorithm, the MSTAR ATR algorithm contains the same general detect, discriminate, and classify components. See the AFRL MSTAR Public web page for more detail on the MSTAR architecture[5].

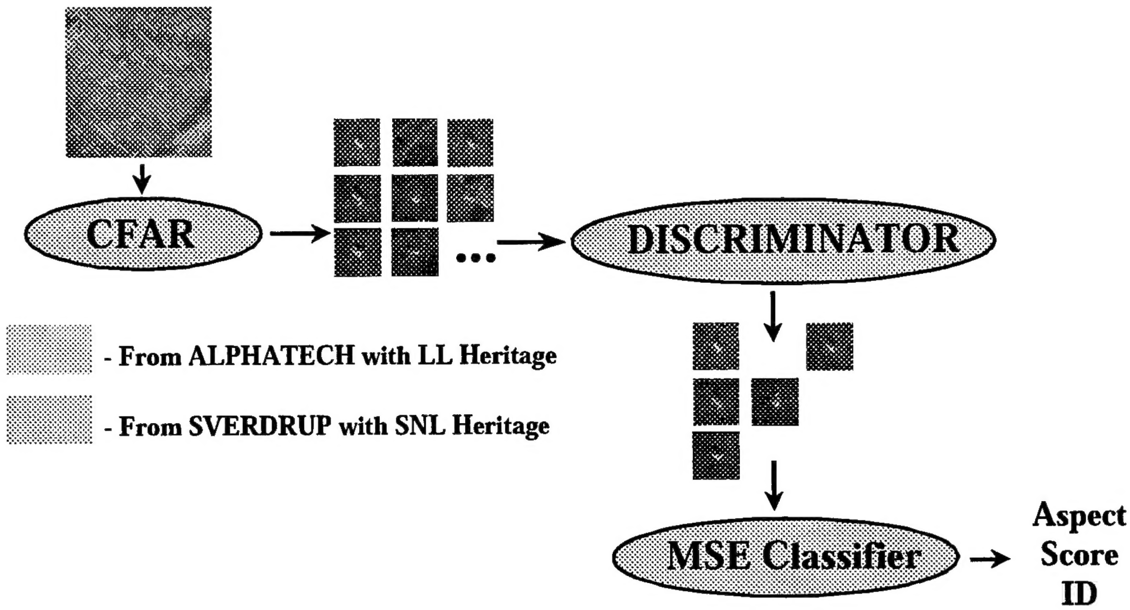


Figure 4. SAR Algorithm Components. The Algorithm consists of three stages a CFAR, Discriminator, and a Classifier.

3.2.2.1 MSE Classifier Metric Definition

In the equations that follow we use the definitions:

$$\dot{x} = \textit{vector form of masked test chip (real-valued)}$$

$$\dot{m} = \textit{vector form of masked mean template (real-valued)}$$

$$\alpha = \textit{scalar scale factor (real-valued)}$$

$$\rho = \frac{\dot{x}^t \dot{m}}{\sqrt{\dot{x}^t \dot{x} \dot{m}^t \dot{m}}} = \textit{correlation between } \dot{x} \textit{ and } \dot{m} \tag{1}$$

The metric implemented by the MSE code is given by

$$MSE = \frac{(\alpha \dot{x} - \dot{m})^t (\alpha \dot{x} - \dot{m})}{\alpha^2 (\dot{x}^t \dot{x})} \tag{2}$$

Where α *is given by*

$$\alpha = \frac{\dot{x}^t \dot{m}}{\dot{x}^t \dot{x}} \tag{3}$$

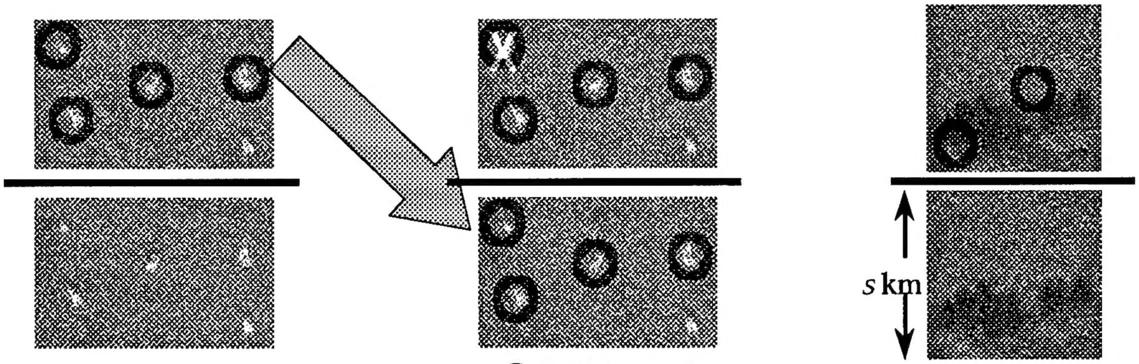
Upon inserting α *in the equation for MSE and simplifying*

we find $$MSE = \frac{1}{\rho^2} - 1 \tag{4}$$

3.2.3 Performance Metrics Used

The third important ingredient necessary for a successful ATR evaluation involves performance metrics. It is important that the performance metrics be clearly defined and well communicated when presenting results. Figure 5 presents three measures

(Pd, Pid, FAR) and graphically depicts their definitions. When reporting these Figures of Merit (FOMs) an associated confidence interval is presented. We do not go into a detailed description here but do in a "sister" publication[6].



$$P_D = \frac{4}{5}$$

Conditiona l

$$P_{ID} = \frac{3}{4}$$

$$FAR = \frac{\#\mathrm{FA's}}{s^2\mathrm{km}^2}$$

Figure 5. Performance Measures, Pd and Pid are measured on target scenes, FAR is measured on clutter scenes.

The MSTAR system goal levels for the three performance measures presented above are:

- Pd > 0.9
- Pid > 0.7
- FAR < 0.05 FAs/km^2

Pd and Pid are measured over target scenes while FAR is measured exclusively over clutter scenes. Pid is conditioned on detection (i.e., # of correctly ID'ed targets / # of detected targets).

4. SUMMARY OF ATR TESTING CONDITIONS

4.1 MSTAR II Goals

A variety of evaluation experiments were designed to test algorithm performance against the MSTAR goals. For success the Pd, Pid, and FAR goal levels need to be met across the following EOCs: target versions, depression (15°-45°), number of classes(10 & 15), articulation (M109 and T72 turret), revetment (2' and 5'), configuration, squint (45°-135°), background, aspect (0-360°), serial number (SN).

4.2 Setting an ATR Operating Point

ATR algorithms can be operated at many operating points. A Receiver Operating Characteristic (ROC) curve is the result of plotting a series of ATR operating points. ROC curves are often used to trade off the probability of target detection (Pd) against a False Alarm Rate (FAR) as shown in Figure 6. Once a ROC curve has been generated, a user can use this curve to set an operating point. In this scenario, the operating point defines the Pd and FAR . Once an operating point is set, a confusion matrix can be computed. The confusion matrix is a matrix of fractions. Each column of the confusion matrix represents a possible system decision category (eg. An ATR trained on 10 target types may make 11 decision types. One for each trained target type and an additional decision type for declaring things to be "non-targets"). Each row of a confusion matrix represents a "type" which was presented to the ATR system. Thus the fraction found in cell$_{i,j}$ in the confusion matrix is the number of times the system reported decision_type$_j$ divided by the number of times test_type$_i$ was passed to the system. The diagonal elements of a confusion matrix represent the fraction of the time that the system made correct decisions. The system Pid can be calculated by averaging the diagonal elements. Therefore a confusion matrix with all one's on the diagonal and zeros in all of the off diagonal elements represents a perfect system identification performance of a Pid = 1.0

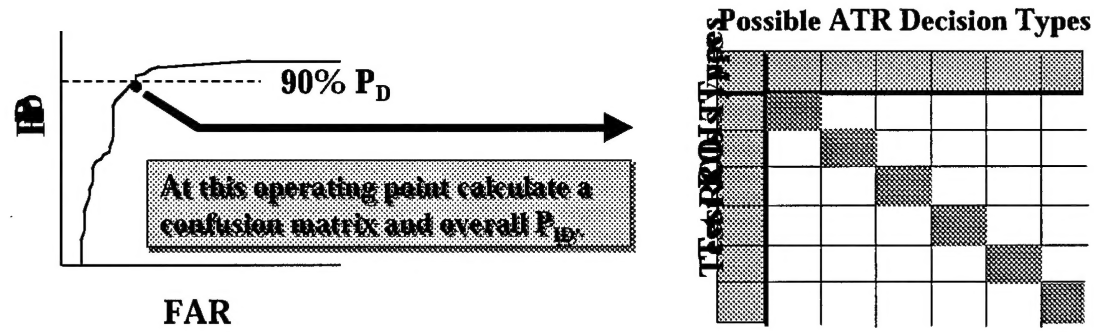


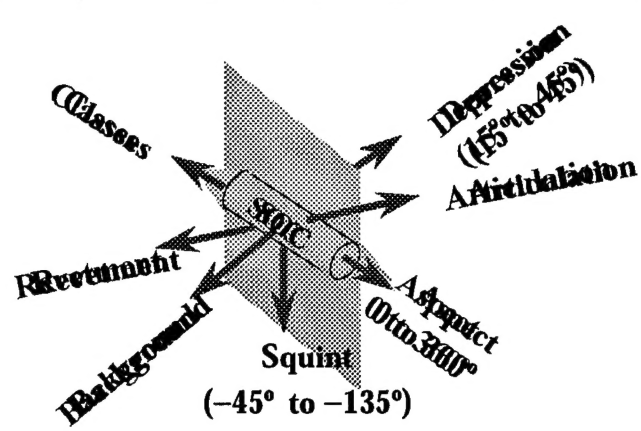
Figure 6. Setting an ATR Operating Point: A ROC Curve and the resulting Confusion Matrix.

For all of the experiments presented in this study the SAR ATR algorithm operating point was set one time **and one time only**. Thirteen experiments will be presented. Each experiment represents algorithm testing over a different target set. Some of these target sets are near the algorithm training conditions and some are not. The ATR algorithm Pd and conditional Pid performance are reported for each experiment. The performance sensitivity as a function of EOCs is then summarized. Note: for the purpose of this study FAR performance sensitivity was not analyzed due to the limited available clutter data.

4.3 Definition of Standard Operating Conditions (nominal conditions)

The Standard Operating Condition (SOC) experiment is defined as the set of operating conditions "very near" training conditions. The SOC test is defined as testing ten target types (M548, M35, Zil131, M2, M113, BTR70, T72, M1, M109, SCUD) across 0°-360° aspect. Seven of the ten target types made use of a single vehicle/serial number during testing. The remaining three target types made use of multiple serial numbers per target type. For the M2s 3 different SNs were tested (MV02GX, MV02FP,T505). For the T72s 3 different SNs were tested (A05, A07, A10). For the M109s 4 SNs were tested (C56, A60, A58, C58). Figure 7 depicts the testing dimensions to be explored. The cylindrical tube represents the standard operating conditions. Each of the labeled vectors off of the tube represent an individual EOC dimension which will be isolated (as best as possible) and tested. The longitudinal axis of the tube spans the "10 target class" and "0°-360° aspect" dimensions. It is represented as a tube as opposed to a line because the SOC test conditions include small deviations in some of the EOC dimensions. The EXP_SOC experiment is defined as test ROIs for the 10 types presented above, sampled across 0°-360° of target aspect, with measured depression angles from 1°-3° away from the ~30° trained depression angle. This EXP_SOC test utilized 630 test ROIs across seventeen different serial numbers. The nominal depression angle for testing was ~30°. For the experiments listed in this paper, unless indicated otherwise, the testing occurred at a depression angle of ~30°.

Figure 7. The Cylindrical tube represents the standard operating conditions. Each vector off of the cylindrical tube



represents an EOC dimension. Each of these EOCs will be isolated and tested individually, some in combinations

4.4 Definition of EOC experiments

To measure SAR ATR extensibility across EOCs, EOC experiments are defined. The measured MSTAR data allows for exploration along many of the individual EOC dimensions. The dimensions of interest are depicted in Figure 7. Each vector off of the cylindrical tube represents an EOC dimension. Each of these EOCs will be isolated and tested individually, some in combinations. In this section, we list seven single EOC experiments (i.e., traversing along an EOC dimension) these experiments are listed in Table 1. In addition to these single EOC experiments, five combination EOC experiments (i.e., the test ROIs are "away" from the SOC conditions in more than one EOC dimension at a time) are defined and are listed in table 2. The table is made up of three columns. The first column presents the experiment name. The Second column lists the target types tested and a brief description. The third column enumerates the "test bins" used in each experiment with a description of each bin along with the number of ROIs used in testing.

Table 1. Single EOC Experiment Descriptions

Experiment Name	Experiment Description	Test Bin Descriptions
EXP_ARTIC	M109 (A60,A58,C58,C56) and T72 (a05,132) were tested. There were 13 discrete turret articulation states from-30° to 90° tested.	1) nominal - 320 SOC ROIs, 0° artic 2) low - 410 ROIs, 9°-40° artic 3) high - 560 ROIs > 40° artic
EXP_REVET	T72(A07,A10) and M109s (A60,A58) were tested. These vehicles were in 2' & 4-5' deep revetments	1) nominal, 180 ROIs, no revet 2) half - 60 ROIs, 2' revet, (a10,a58) 3) full - 60 ROIs, 4-5' revet,(a07,a60)
EXP_COLL	T72, M109 and Zil-131s were tested. All training was done with collection 2 vehicles, testing was done across all three MSTAR collections.	1) col2_1- trained in collection 2 tested in collection 1, 60 ROIs 2) col2_2 -trained in collection2 tested in collection2, 270 ROIs 3) col2_3 trained in collection 2 tested in collection 3, 120 ROIs
EXP_DEP	9 target types were tested (the 10 class problem listed in section 4.2 minus the SCUD). The ATR was trained at ~30° depression and tested at a ~15° and ~45° depression angles.	1) nominal, 300 ROIs 2) ~15° depression angle, 270 ROIs 3) ~45° depression angle, 270 ROIs
EXP_15	10 SOC target types tested against a 10 class ATR. Then 5 additional classes were added to the ATR (making it a 15 class classifier). And the 10 SOC target types were tested against a 15 class problem	1) trained with 10 classes tested with 10 classes, 630 ROIs 2) trained with 15 classes tested with 10 classes, 630 ROIs
EXP_SU	M2s and T72s Tested. Testing Sequestered ROIs vs untrained Unsequestered ROIs.	1) Tested on Sequestered, 120 ROIs 2) Tested Unsequestered, 180 ROIs
EXP_SN	M2s & T72s tested. Measuring performance difference between trained SNs but untrained ROIs against untrained SNs and untrained ROIs	1) ROIs from trained SNs but not the training ROIs, 60 ROIs 2) ROIs from untrained SNs, 180 ROIs

Table 2. The Combination EOC experiment Descriptions

Experiment Name	Experiment Description	Test Bin Descriptions
EXP_ART_RET	T72s tested in their nominal state versus T72s in a combination of low or high articulation coupled with being in 2' or 5' revetments.	1) Nominal, 90 ROIs 2) ART+RET, 140 ROIs
EXP_ART_DEP	T72s, M109s tested. At their nominal state versus in a combination of low or high articulation coupled with being 15°s away from the trained depression angle.	1) Nominal, 180 ROIs 2) ATR+DEP, 290 ROIs
EXP_RET_DEP	T72s tested in their nominal state versus T72s in a combination 2' or 5' revetments coupled with being 15°s away from the trained depression angle.	1) Nominal, 90 ROIs 2) RET+DEP, 110 ROIs
EXP_SQU_COL	T72, M109 and Zil-131s were tested. All training was done with collection 2 vehicles. All testing was done using collection 3 imagery.	1) Broadside squint -100°-80°s, 120 ROIs 2) Aft squint,-135°-125°s, 120 ROIs 3) Forward squint,-55°-45°s, 120 ROIs
EXP_SQU_DEP_COL	T72, M109 and Zil-131s were tested. All training was done with collection 2 vehicles. All testing was done using collection 3 imagery coupled with being 15°s away from the trained depression angle	1) Nominal, 120 ROIs 2) SQU+DEP_COL, 240 ROIs

Results are presented in Section 5 for the EXP_SOC experiment; each of the 7 single EOC experiments listed in Table 1; and for each of the five combination EOC experiments described in Table 2.

5. EXPERIMENTAL RESULTS

All of the experiments listed in sections 4.3 and 4.4 were run through the baseline SAR ATR algorithm described in section 3.2.2. A single ATR operating point was set according to the procedure laid out in section 4.2. The operating point was set to achieve a Pd = 0.9 using the union of the test ROIs found in each of the twelve EOC experiments described above. An individual probability of detection (Pd) and probability of correct identification (Pid) were then computed at this operating point.

5.1 ATR Pd Results

Figure 8 depicts the SAR ATR algorithm's Pd performance sensitivity as a function of the thirteen experiments. Eight of the thirteen experiments have near perfect target detection performance. Three of the experiments have detection performance at the goal level of Pd = 0.90. Two experiments, EXP_DEP and EXP_SQU_DEP_COL experienced detection performance degradation below that of the goal level at approximately Pd = 0.80. All in all this SAR ATR algorithm while only trained at conditions near the EXP_SOC test performed quite well against the MSTAR II EOCs in a Pd sense. We don't find this too surprising since all of the EOCs except the revetment EOC were designed to result in almost no information loss from the target signature. The fact that the revetments caused little detection difficulty to this algorithm was pleasantly surprising. Upon review of the SAR signatures (from a ~30° depression angle) of the revetted ROIs, the man made revetments caused little obscuration or layover effects. This was mostly due to the fact that the targets had a few feet of space between the walls of the revetments and the target sides.

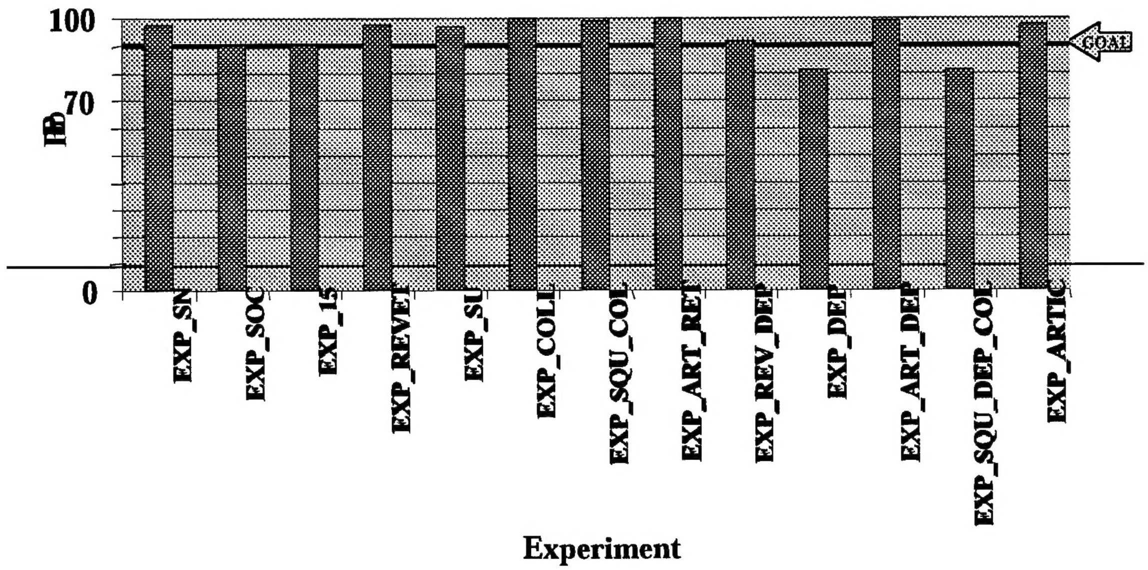


Figure 8. This Bar Chart depicts the Probability of Detection Performance versus Experiment for the baseline SAR ATR algorithm under test. Note: all but two experiments are at or above the goal level of Pd = 0.9.

5.2 ATR Pid Results

Figure 9 depicts the Pid performance for each of the experiments tested. They are sorted into three groups: those at or above the goal level, those below the goal level but greater than 50%, and those below 50%. The confusion matrix in Figure 10 is a result of combining the 12 EOC experiments into a single experiment. The resulting overall Pid = 0.50

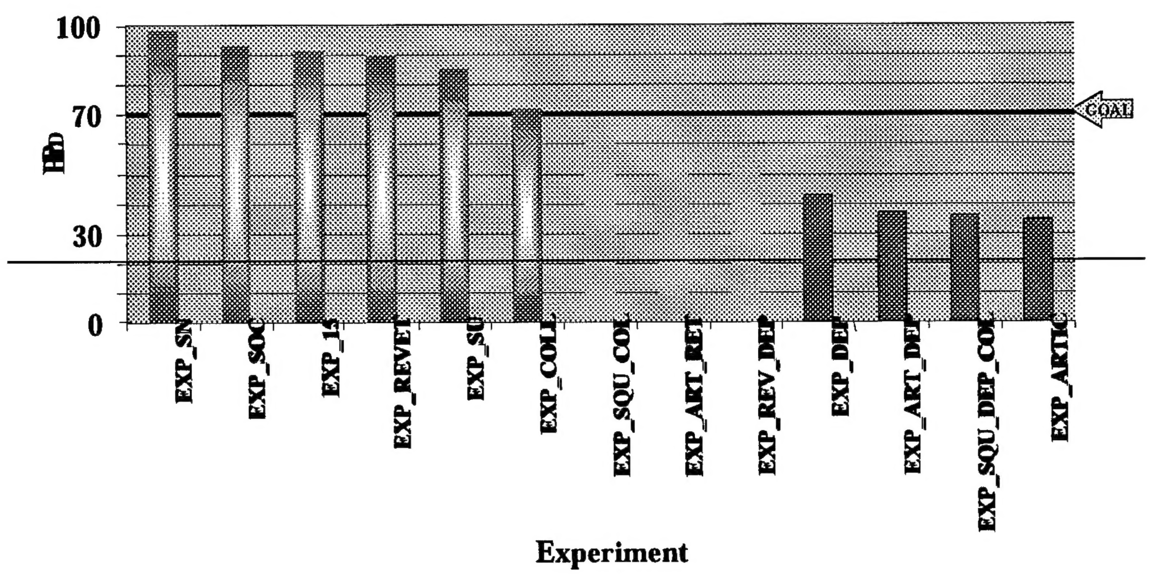


Figure 9. This Bar Chart depicts the Probability of Correct Identification (Pid) Performance versus Experiment for the baseline SAR ATR algorithm under test. Note: Pid performance is very sensitive to EOC conditions. Six of the thirteen experiments were at or above the Pid goal level of 0.70. Seven of the experiments were measurably below the goal level and four of the EOC conditions tested resulted Pid's < 0.50.

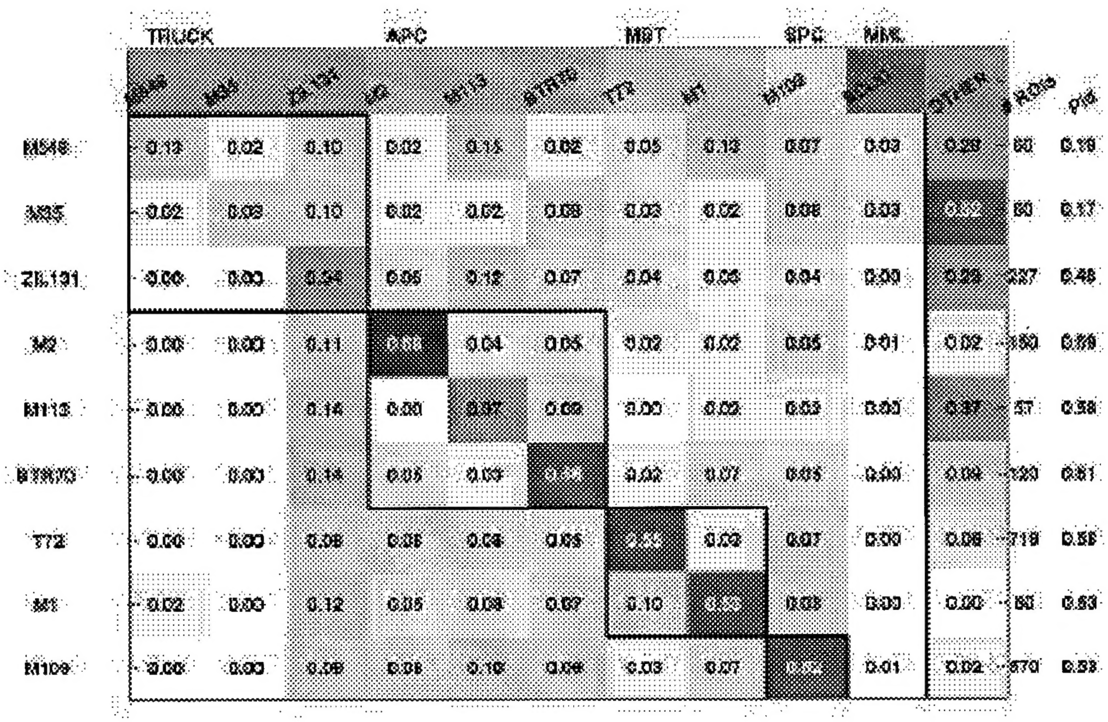


Figure 10. A confusion matrix representing the Baseline SAR ATR identification performance over the 12 EOC experiments presented in Tables 1 and 2. The overall Pid calculated by averaging the performance across the target types (i.e., the average of the diagonal of the confusion matrix) is Pid = 0.50.

5.2.1 Tests resulting in Pid performance at or above goal level

- EXP_SN The serial number test, EXP_SN, when all EOC dimensions are controlled as much as possible and the Serial Number is the difference from the training and the testing conditions, the baseline ATR's Pid performance was nearly perfect. We cannot conclude from this test that: training on SN X and testing on SN Y (each from the same target class) for all of the X's and Y's in the real world will result in sustained Pid performance as seen in this test. What we can conclude from this test is that if controlling for factors like target version variant differences, target configuration differences, target damage, ... then difference in SN is not the Pid performance driver.
- EXP_SOC When testing an ATR very near the training conditions one would expect excellent Pid performance. It was found to be true with this test.
- EXP_15 When adding five additional potential confuser classes to this ATR, the Pid performance when testing SOC ROIs is not effected
- EXP_REVET The 2' and 5' deep man made revetments used in the MSTAR data collection had very little effect on the SAR signature of the MSTAR targets when viewed from a 30° depression angle and the associated Pid was minimally impacted.
- EXP_SU Testing occurred across ROIs which were not Sequestered but were not trained on by the ATR algorithm and the Pid performance was minimally impacted. This isn't a big surprise this test was designed to measure if in fact the ATR algorithm was over trained on attributes found in the Unsequestered data.
- EXP_COLL When all EOCs dimensions were controlled (with the best of our ability) and the difference was a Collection EOC the ATR algorithm Pid performance dropped to the goal level of Pid = 0.70.

5.2.2 Tests resulting in Pid performance below goal level but above Pid = 0.50

- EXP_SQU_COL The combination of +/- 45° squint from the training condition and test ROIs from a collection different from the training condition resulted in Pid of approximately 0.60 on this MSE based SAR ATR algorithm. Note: all of the training and testing of this algorithm was done in the ground plane. Since the squint angle of the radar is an acquisition parameter it is known and was accounted for when projecting the imagery from the image slant plane into the ground plane.
- EXP_ART_RET Revetments by themselves did not cause Pid performance degradation with the baseline algorithm. However when coupled with the Articulation EOC the Pid ~= 0.60. Note: when articulation was treated as a single EOC, the Pid ~=0.35.
- EXP_RET_DEP Revetments coupled with the depression EOC resulted in a Pid performance of ~=0.60.

5.2.3 Tests resulting in Pid performance below 0.50

- EXP_DEP Depression (i.e., test ROIs +/- 15°s away from the training depression angles) as a single EOC condition resulted in severe Pid performance degradation. The resulting Pid ~= 0.40.
- EXP_ART_DEP Articulation + Depression EOCs when coupled, resulted in a Pid ~= 0.35. That is, approximately 1 out of 3 of the T72s or M109s tested in this experiment were correctly identified by the ATR algorithm.
- EXP_SQU_DEP_COL When the test condition was away from the training condition in three dimensions (Squint, Depression, Collection) the resultant Pid ~= 0.35.
- EXP_ARTIC M109s and T72s were tested at 13 different turret articulation states (-30°-90°). The algorithm was trained at a single articulation state with the turret facing straight ahead (0° articulation). This MSE SAR ATR algorithm had a Pid ~=0.35.

6. SUMMARY

Testing a SAR Automatic Target Recognition (ATR) algorithm at or very near its training conditions yielded near perfect performance in both a Pid and Pd sense. Twelve experiments were set up to measure the algorithm's Pid and Pd sensitivity to Extended Operating Conditions. All of these EOCs (with the exception of revetments) were designed to have little or no impact on the target's SAR signature with respect to "information loss". **As a result the ATR algorithm's ability to achieve high probability of detection (Pd) across these EOC states was not impacted.** As it turned out, the revetment test cases had virtually no impact on the ATR's ability to perform detection or identification. In fact, when the revetment was coupled with either the Articulation or Depression EOC, the Pid of the combination EOC was greater than the Pid for the Articulation or Depression EOC when tested in isolation. This "revetment helping" phenomena shouldn't be given too much weight as the confidence intervals surrounding these Pid's did slightly overlap. **In general, the SAR ATR Pid performance was very sensitive to the EOCs tested.**

7. ACKNOWLEDGEMENTS

This paper is the result of contributions by the entire SvT/AFRL Performance Evaluation team. We would especially like to thank Bob Kotz, Jason Johnson, Ron Dilsavor of Sverdrup Technology and Bill Irving of Alphatech for constructing, hosting and training the SAR ATR algorithm used in this paper. We would like to thank Lannie Hudson and Andy Morrison from Sverdrup Technology, for running the experiments and conducting results analysis. Last but not least, we would like to thank Jason Johnson and Mike O'Connor for assisting in the manuscript preparation.

8. REFERENCES

1. DARPA's MSTAR public web site http://maco.dc.isx.com/iso/battle/mstar.html
2. E.R. Keydel, S.W. Lee, J.T. Moore, "MSTAR extended operating conditions: a tutorial", Proc SPIE, Vol. 2757, pp.228-242, Algorithms for synthetic Aperture Radar Imagery IV, April 1997
3. T. D. Ross, L. Westerkamp, E. Zelnio, T. Burns, "Extensibility and other model-based ATR Evaluation Concepts" Proc. SPIE, Vol. 3070, pp.213-222, Algorithms for synthetic Aperture Radar Imagery IV, April 1997
4. AFRL Data teams' public web site http://www.mbvlab.wpafb.af.mil/public/MBVDATA/
5. AFRL's MSTAR public web site http://134.131.123.60/
6. T.D. Ross, V. Velton, J. Mossing, S. Worrell, M. Bryant, "Standard SAR ATR Evaluation Experiments using the MSTAR Public Release Data Set", Proc. SPIE, vol 3370, SPIE'98 Algorithms for synthetic Aperture Radar Imagery V, April 1998